CGIAR Research Program on
Climate Change, Agriculture and Food Security (CCAFS)

Archiving & Sharing
Video Transcript

October 2013

These materials were produced for and with funding from the Climate Change Agriculture and Food Security Research Program of the Consultative Group on International Agricultural Research (CGIAR).

Archiving & Sharing

This short video considers some of the benefits of archiving research data and highlights ethical aspects such as gaining consent from respondents and confidentiality of personal data.

Why should I archive my Data?

In the past, and still today in some cases, researchers have been reluctant to share their data. This might be because they fear someone else might get the credit for the work they have done or perhaps they want more time to do their own analysis.

However, most funding bodies are now viewing research data as a public resource and insisting that data are put into the public domain in a timely fashion.

Advantages of sharing data

There are many advantages to sharing data including:

- Encouraging scientific enquiry and debate
- Promoting innovation and potential new uses of the data
- Possible new collaborations between researchers
- Maximising transparency and accountability
- Encouraging the improvement and validation of research methods
- Eliminating the need to collect the same data again thus reducing time and costs
- Increasing the impact and visibility of the research.

Principles

General principles for archiving data might include:

- Publicly funded research data are a public good which should be made openly available in a timely and responsible manner
- Sufficient metadata should be recorded and made available to enable others to understand the research and the potential re-use of the data
- Researchers should have a limited period of privileged use of the data collected to enable them to publish the results of their research – the time might vary but is expected to be no longer than 2 years
- All users of the data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed

Anonymity

You should not be able to identify individuals from archived data. With quantitative data anonymisation is relatively straight-forward – you might for instance remove direct identifiers – this would include names, addresses, phone numbers, etc. There are not usually necessary for secondary research but are collecting for checking purposes or to enable follow-up. These variables should be replaced by a code in the data.

You might aggregate or reduce the precision of a variable – for example rather than the full birthdate of an individual you might just record the month and year.

If the values for an individual respondent are unusual within the wider group researched you could collapse unusually large or small values into a single code. For example a top code of "50 hectares or more" could be applied to land ownership.

With qualitative data such as interview transcriptions removing identifiers could distort the data. One suggestion is to use pseudonyms, replacement terms or vaguer descriptions – for example instead of the respondent's real name (Cathy for instance) you could use a pseudonym replacing every occurrence of "Cathy" with "Jane" in all the transcripts; instead of "Station Hill Primary School" you could use "A primary school". When making these changes keep a copy of the original data for internal use and keep a log of all replacements used.

If you have audio-visual data to anonymise this is much more of a challenge – if this is likely to be an issue it is better to obtain the participant's consent to use and share the data unaltered.

Consent

Informed consent is an ethical requirement for a lot of research. When gaining consent you must make provision for sharing & archiving the data.

Respondents should be given information on the purpose of the research and what is involved if they agree to participate. Any benefits or risks to them as individuals must be made clear. They should also be told how you are intending to use the data, how it is to be stored and how you are planning on ensuring confidentiality of their personal data.

If you are intending to use audio-visual data such as recordings or photos you should state this clearly – some might agree to complete a questionnaire but might not want photos of themselves to be made public.

If your research involves working with children then you must gain consent from the parent or guardian as well as from the child (assuming the child is able to understand what is being asked of them).

What should I archive?

Of course when we archive data it is not just the data file that we archive. Many data files are of limited use without the accompanying documentation. At the minimum your archive should include

- The activity protocol so others can clearly see the focus of your research
- The data management plan to show the steps you intended to follow to ensure high quality data
- If you have used a data entry system this should also be included
- The fieldworker manual will detail the procedures used to collect the data
- Include the blank questionnaires – adding variable names to the questionnaire would be useful
- The data quality report would highlight any problem areas in the data and give suggestions for its use.
- The metadata document is used to describe the data
- Etc.

Summary

So to summarise there are clear benefits to archiving and sharing but there are also responsibilities - you must ensure the confidentiality of your respondents - ensure you have informed consent from respondents and your data have been anonymised.

Think of data sharing as a two-way process – if you are not willing to share your data you cannot expect others to share theirs with you. The section of the support pack on Data Ownership includes further resources on this issue.

Appendix I – CCAFS Data Management Support Pack

This document is part of the CCAFS Data Management Support Pack produced by the Statistical Services Centre, University of Reading, UK. The following materials are available in the pack:

0. Data Management Strategy
    a. CCAFS Data Management Strategy
1. Research Protocols
    a. Writing Research Protocols – a statistical perspective
    b. Preparation of Research Protocols – Good Practice Case Study
    c. What is a Research Protocol, and how to use one (Video & Transcript)
    d. Details of what a Research Protocol should contain (Video & Transcript)
2. Data Management Policies & Plans
    a. Creating a Data Management Plan
    b. Data Management Plan (Video & Transcript)
    c. Example Data Management Activity Plan
    d. Example Consent Form
3. Budgeting & Planning
    a. Budgeting & Planning for Data Management
    b. ToR Data Support Staff
    c. Budgeting & Planning (Video & Transcript)
4. Data Ownership
    a. Data Ownership and Authorship
    b. Template – Data Ownership Agreement
    c. CCAFS Data Ownership & Sharing Agreement
    d. Data Ownership & Authorship (Video & Transcript)
5. Data & Document Storage
    a. Creating and Using a DDS
    b. DDS Introduction – (Video & Transcript)
    c. DDS Organisation – (Video & Transcript)
    d. DDS Ownership – (Video & Transcript)
    e. Introduction to Dropbox – (Video & Transcript)
6. Archiving & Sharing
    a. Archiving & Sharing Data
    b. Data and Documents to Submit for Archiving – a checklist
    c. MetaData
    d. Archiving & Sharing (Video & Transcript)
    e. Metadata (Video & Transcript)
    f. CCAFS HBS Questionnaire
    g. CCAFS HHS Code Book
    h. CCAFS Training Manual for Field Supervisors

7. CCAFS Data Portals
    a. Portals for CCAFS Outputs
    b. AgTrials Summary
    c. CCAFS-Climate Summary
    d. DSpace Introduction
    e. Introduction to Dataverse (Video & Transcript)
    f. Creating a Dataverse (Video & Transcript)
    g. Dataverse Study Catalogue
    h. CCAFS Dataverse (Video & Transcript)
8. Data Quality & Organisation
    a. Data Quality Assurance
    b. Guidance for handling different types of Data
    c. Transition from Raw to Primary Data
    d. Data Quality Assurance (Video & Transcript)
    e. Guidance for handling different types of data (Video & Transcript)
    f. Transition from Raw to Primary Data (Video & Transcript)